Power System Technology
ISSN:1000-3673

Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

Detecting and Preventing Sql Injection Attacks with Machine Learning Predictive Analytics

M. Ravi Chandra[1], V.Ramanjaneyulu[2], O.Bhavya Sri[3],B.Nagarjun Singh[4], M.Paul Sundar Singh[5]

[*1](PG student, Dept of CSE, Mandava Institute of Engineering and Technology- Jaggaiahpet-521175,India.)

[*2] (Asst Professor, Dept of CSE, Santhiram Engineering College- Nandyal-518501,India.)

[*3,4,5] (Asst Professor, Dept of CSE, Mandava Institute of Engineering and Technology- Jaggaiahpet-521175,India.)

Abstract:-

The storage of the enormous volume of large data that is exchanged over the Internet from cloud-hosted web apps to Internet of Things (IoT) smart devices depends heavily on the back-end database. On weak web applications, the Structured Query Language (SQL) Injection Attack (SQLIA) is still the go-to attack for hackers looking to steal private information from databases with potentially harmful outcomes. The current solutions, which mostly use signature techniques, were developed prior to the latest big data mining issues and, as a result, lack the functionality and capacity to handle new signatures that are hidden in web requests. Predictive analytics using alternative machine learning (ML) offers a scalable and useful way to mine massive data for SQLIA detection and mitigation. Unfortunately, a well-known problem in SQLIA research is the absence of readily available robust corpuses or data sets having patterns and historical data items to train a classifier. In this project, we investigate the creation of a data set that includes extraction from well-known attack patterns, such as SQL tokens and injection point symbols. In order to demonstrate vast amounts of learning data, we also construct a web application as a test case that anticipates dictionary word lists as vector variables. For supervised learning, the data set has been pre-processed, labeled, and feature hashed. In order to prevent malicious web requests from reaching the protected back-end database, the trained classifier will be deployed as a web service that is used in a custom dot NET application that implements a web proxy Application Programming Interface (API). This will allow the classifier to intercept and accurately predict SQLIA in web requests. With empirical assessments shown in the Confusion Matrix (CM) and Receiver Operating Curve (ROC), this project shows a complete proof of concept implementation of an ML predictive analytics and deployment of the resulting web service that accurately predicts and prevents SQLIA

Keywords: Hybrid energy storage system, Battery, Supercapacitor, Electric vehicles & Regenerative braking

Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

SQLIA, SQLIA analytics, SQL Injection, SQLIA big data, SQLIA hashing.

1. Introduction

A computer programming language called a query language is used to access and modify data stored in databases. In order to carry out tasks like adding, editing, removing, and retrieving data, it enables users to interact with the database management system (DBMS).

Fundamentally, a query language lets you create statements that make it easier to query, modify, and retrieve data from a database. It serves as a mediator, converting queries from people into commands that databases can comprehend. Users may communicate and get the information they require thanks to this translation, even if they lack in-depth technical understanding of the database architecture or storage systems.

In reality, distinct query languages are designed for particular database types and uses. Each has a specific use, ranging from the popular SQL for relational databases to SPARQL, which is made for searching RDF data on the Semantic Web.These languages enable users to quickly sort through vast amounts of data. This supports downstream data analytics initiatives, such as identifying patterns and promoting well-informed choices.

A vulnerability in the database layer of an application is exploited using an attack technique known as SQL injection. Hackers employ injections to get unauthorized access to the structure, underlying data, and database management system. An attacker can use SQL injection to deliver malicious code to the backend database after inserting it into a poorly designed application. Following that, the malicious data produces inappropriate actions or database query results. If the right circumstances are met, an attacker can use a SQL Injection vulnerability to bypass authorization and authentication checks in a web application and get access to an entire database's contents.

The insertion, editing, and deletion of records in a database by SQL Injection may have an effect on data integrity. To this extent, SQL Injection can give an opponent unauthorized access to confidential data. The SQL injection technique is used to attack data-driven systems by inserting malicious SQL statements into an entry field for execution (e.g., to dump the database contents to the attacker). An application's software weakness must be exploited by SQL injection.

II. LITERATURE SURVEY

A. Enhancing SVM-Based SQLIA Prediction with Comprehensive Data Sets and Pre-Processing Techniques

The effectiveness of text pre-processing techniques and the caliber of data sets are essential for improving SVM-based SQLIA prediction. This section examines techniques for producing trustworthy training data, incorporating a range of patterns, and improving text pre-processing in order to enhance the performance of SVM classifiers in predicting

Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

SQLIA.Data augmentation techniques can be used to further improve the training data for SVM-based SQLIA prediction. techniques such as oversampling minority classes and generating synthetic data utilizing SMOTE (Synthetic Minority Over-sampling Technique).

B. Challenges in Data Engineering for SVM-Based SQLIA Mitigation

Data engineering is one of the main obstacles to using Support Vector Machine (SVM) machine learning for SQL Injection Attack (SQLIA) mitigation.
Although SVM has the potential to strengthen security protocols, its efficacy is mostly dependent on the caliber of feature extraction and data preparation. Current SVM-based methods often struggle to interpret text data effectively, which is a crucial component of SQLIA detection. The absence of thorough text preparation methods frequently results in an inability to precisely identify the subtle patterns suggestive of SQL injection attempts. Because of this, these flaws make online applications more vulnerable to SQLIA assaults by reducing the efficiency of SVM classifiers in thwarting them. A focused effort must be made to enhance data engineering practices in order to get past these challenges, with a focus on trustworthy text preparation methods created expressly for the challenges of SQL injection detection. By enhancing the quality of input data and optimizing text preparation methods, SVM-based systems can be more suited to identify and mitigate SQLIA, strengthening the security posture of web applications against malicious exploitation.

C. Gap in ML Application for Predicting SQLIA in Big Data Contexts

The use of machine learning (ML) in the big data area to predict SQLIA has not received much attention up to this point. The focus on patterns and text pre-processing in the Multi-Aspect Multi-Layer (MAML) architecture remains untapped, despite the potential for significant improvements in prediction accuracy.

III. EXISTING SYSTEM

The SQLIA keywords are also in plain text, and the existing implementations for the syntax of the SQL language closely resemble plain English. Consequently, a supervised learning model trained using both known historical attack signatures and safe online request patterns makes the SQLIA problem in a large data setting a credible choice for predictive analytics. While legitimate web requests will take the shape of expected data from the application, attack signatures at injection locations will contain patterns of SQL tokens and symbols as SQLIA positive. In order to train a classifier, we construct a predictive analytics web application using large amounts of learning data. The learning data are labeled vector matrices, which are features of both SQL tokens (SQLIA positive) and dictionary word list patterns (SQLIA negative).

Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

IV. PROPOSED SYSTEM

To train a classifier in the suggested method, we create a prophetic analytics web operation using a large amount of learning data. The literacy data are labeled vector matrix, or features of both SQL commemorative (SQLIA positive) and dictionary word list (SQLIA negative) patterns. By using a realistic data set that experiences point mincing, this project's benefits include training a supervised literacy model that uses the Support Vector Machine (SVM) method to predict SQLIA directly, preventing malicious web requests from getting to the target back-end database. Additionally, it provides a big data internet environment for SQLIA discovery. Additionally, this project provides proof of concept for a functional prototype that uses Microsoft Azure Machine Learning (MAML)-enforced Two-Class Support Vector Machine (TCSVM) machine learning techniques to predict SQLIA. The empirical evaluation of Receiver Operating Curve (ROC) also focuses on this methodology.

V. SYSTEM ARCHITECTURE

A web service containing the trained model is made available. For continuous SQLIA detection and prevention, the web service is called in NETSQLIA, a specially designed dot NET program for this study. Retraining the classifier to adapt to a new environment requires the administrator or system expert to provide the data engineering or text pre-processing module with a new rule that matches the patterns in the new data set. This is crucial for deployment in every new domain.

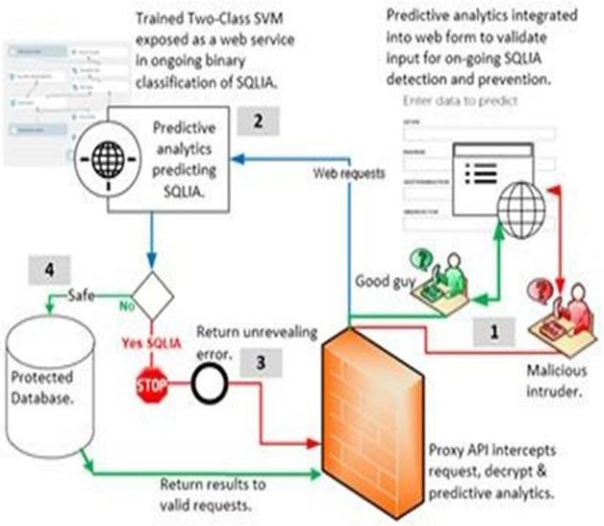


Figure 1:A custom application is consuming the trained SVM web service for ongoing SQLIA detection and prevention

Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

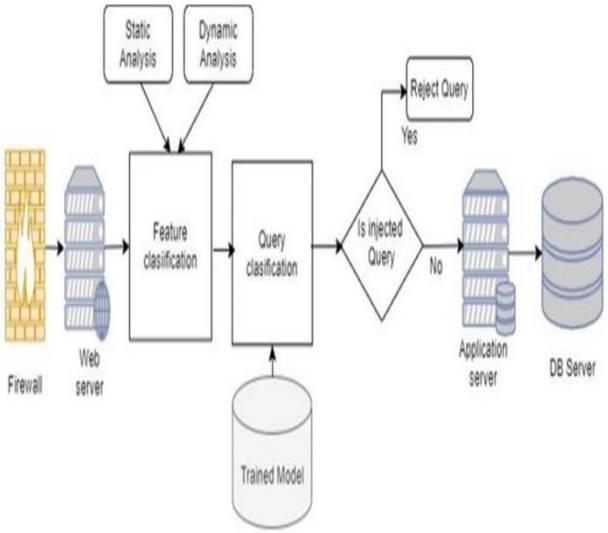


Figure 2: System Architecture

VI. STEPS TO IMPLEMENT PROPOSED MODEL

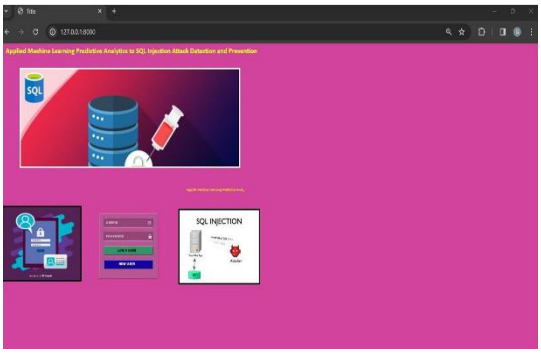
1) Test Preprocessing: In this step, a model is trained for continuous detection and prevention using R scripting and regular expression pattern matching. Patterns of both legitimate and invalid requests are used to increase the data set items in a real-world domain application. Following text pre-processing of the parsing data set for patterns, duplication, normalization to lower cases, and the removal of missing words, there were 362,603-row items.To ensure an even distribution of row items (records), the data set is sampled. The Synthetic Minority Over-Sampling Technique (SMOTE) was used to rectify the unbalanced data set (most negatives over positives). This resulted in 725206 items, divided equally between 362,603-row items of attack/respondent (positives) and 362,603-row items of non-attack/non-respondent (negatives). These steps increase the precision and recall of the trained model.

2) SQL Injection Point: The best way to intercept requests coming from any injection technique is to use a web proxy API. Any Web page form, such as the login screen, can serve as the source of an injection mechanism. Other methods include second-order injection, which hides a Trojan horse for the attack at a later time, exploiting web-enabled server variables to access the back-end database, and using cookies that store state information to access the back-end database without authorization. The SQLIA types—which include tautology, invalid/logical incorrect, union, piggybacking, store procedure, time-based, and alternate encoding obfuscation—are methods an attacker might use at injection locations in any combination to execute an attack. During labeling, SQLIA types offer an extract for the SQLIA positive in data set items.

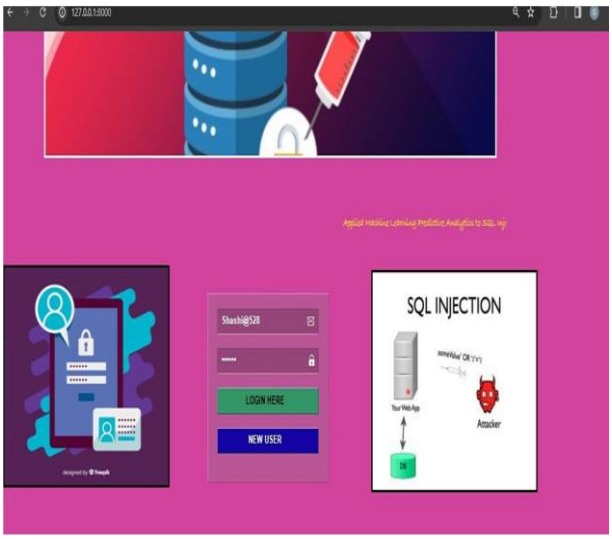
Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

3) Proxy Filters: The ability to decrypt obfuscated internet data for in-depth analysis is a benefit of the third technique, proxy filters, which intercept web requests at a proxy for SQLIA detection and prevention. For SQL syntax sequence alignment, we suggest a SQL parsing tree that combines a proxy and a SQL parser tree. In order to backhaul web requests for predictive analytics of incoming web requests for SQLIA negatives and positives, the strategy suggested in this research makes use of proxy API.

4) Classifying Attacks: In this section, we evaluate SVM's classification capabilities against those of other well-known machine learning algorithms. A number of well-known classification algorithms have been chosen. We try to optimize the performance of each method by utilizing multiple sets of parameters. SVM methods are used for malware bag-of-words weighting classification.

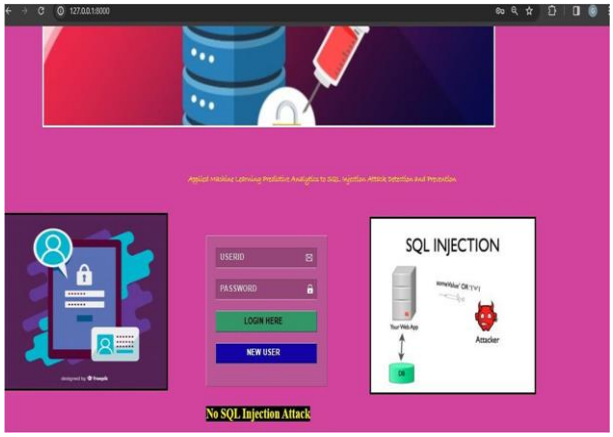


Fig(d): Output Screen

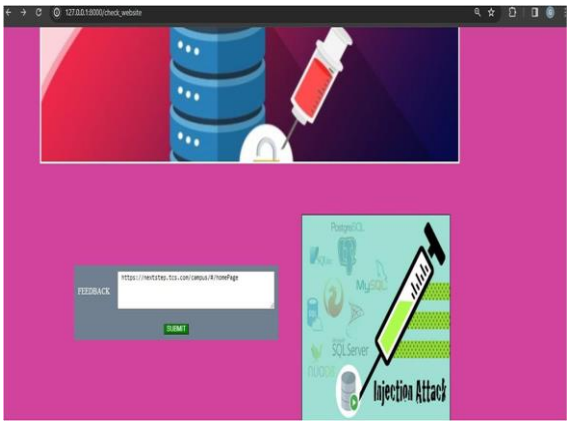


Fig(e): User Login Page

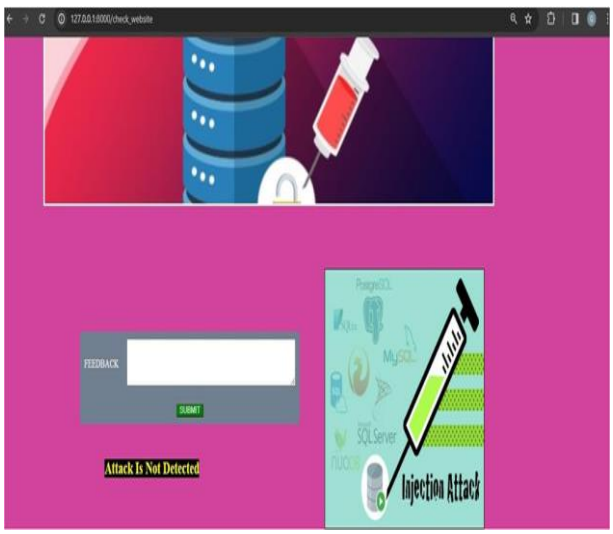
Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025



Fig(f): No SQL Injection Attack



Fig(h): Choose Website



Fig(i): Feedback about Website

Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

VIII. CONCLUSION

In this project, we showed how to use predictive analytics for SQLIA detection and prevention in a big data setting, with great results that are empirically assessed in the ROC graph and confusion matrix. When comparing this study to previous research, the methodology suggested here performs well in a massive data setting, which is uncommon in previous SQLIA research to the best of our knowledge. In order to detect and classify the various SQLIA kinds as predicted, future work will use multi-class classifiers.

IX.REFERENCES:

[1] Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm FS Mahammad, P Bhaskar, A Prudvi, NY Reddy, PJ Reddy journal of algebraic statistics 13 (3), 40-45.

[2] Machine Learning Based Predictive Model for Closed Loop Air Filtering SystemP Bhaskar, FS Mahammad, AH Kumar, DR Kumar, SMA Khadar, ... Journal of Algebraic Statistics 13 (3), 609-616

[3] Devi, M. S., Mahammad, F. S., Bhavana, D., Sukanya, D., Thanusha, T. S., Chandrakala, M., & Swathi, P. V. (2022).” Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection.” Journal of Algebraic Statistics, 13(3), 112-117.

[4] Devi, M. M. S., & Gangadhar, M. Y. (2012).” A comparative Study of Classification Algorithm for Printed Telugu Character Recognition.” International Journal of Electronics Communication and Computer Engineering, 3(3), 633-641.

[5] Devi, M. S., Meghana, A. I., Susmitha, M., Mounika, G., Vineela, G., & Padmavathi, M. MISSING CHILD IDENTIFICATION SYSTEM USING DEEP LEARNING.

[6] Kumar, M. S., Harika, A., Sushama, C., & Neelima, P. (2022). Automated Extraction of Non-Functional Requirements From Text Files: A Supervised Learning Approach. Handbook of Intelligent Computing and Optimization for Sustainable Development, 149-170.

[7] Devi, M. S., Poojitha, M., Sucharitha, R., Keerthi, K., Manideepika, P., & Vasudha, C. Extracting and Analyzing Features in Natural Language Processing for Deep Learning with English Language.

[8] B.Krishna Naga Deepthi, Dr.M.V.Subramanyam,” Analysis And Optimization Of Power And Area Of Domino Full Adder And Its Applications”, Iosr Journal Of Electronics And Communication Engineering, Vol.10,No.3,Pp.55-63,2015.

[9] Y.Murali Mohan Babu, Dr.M.V.Subramanyam,M.N. Giri Prasad,” A New Approach For Sar Image Denoising”, International Journal Of Electrical And Computer Engineering, Vol.5,No.5,Pp.984-991,2015. (Scopus Indexed)

Power System Technology
ISSN:1000-3673
Received: 16-04-2025 Revised: 05-05-2025 Accepted: 22-07-2025

[10] Ch.Nagaraju, Dr.Anil Kumar Sharma, Dr.M.V.Subramanyam," A Review On Ber Performance Analysis And Papr Mitigation In Mimo Ofdm Systems", International Journal Of Engineering Technology And Computer Research, Vol.3,No.3,Pp.237-238, June, 2015.

[11] D.Lakshmaiah, Dr.M.Subramanyam, Dr.K.Satya Prasad," Design Of Low Power 4- Bit Cmos Braun Multiplier Based On Threshold Voltage Techniques", Global JOURNAL OF RESEARCH IN ENGINEERING, VOL.14(9),PP.1125-1131,2014.

[12] R Sumalatha, Dr.M.Subramanyam, "Image Denoising Using Spatial Adaptive Mask Filter", Ieee International Conference On Electrical, Electronics, Signals, Communication &Amp; Optimization (Eesco-2015), Organized Byvignans Institute Of Information Technology, Vishakapatnam, 24 Th To 26th January 2015. (Scopus Indexed)

[13] P.Balamurali Krishna, Dr.M.V.Subramanyam, Dr.K.Satya Prasad, "Hybrid Genetic Optimization To Mitigate Starvation In Wireless Mesh Networks", Indian Journal Of Science And Technology,Vol.8,No.23,2015. (Scopus Indexed)